7

From sherds to blocks: statistics and the archaeological sample

Sarah Vaughan
University College London

Dave Guppy
Computer Centre, University College London

7.1 Introduction

In the not-too-distant past archaeology underwent a fundamental change from artifact acquisition and description to scientific retrieval and data analysis in the context of theory. One important aspect of this change was the imposition of new demands on systems of study and classification of material remains, particularly in the field of ceramics; a field which still constitutes the class of artifact found in best preservation and greatest abundance at most excavations. An increasing awareness of ancient technology combined with more frequent application of laboratory analyses meant that both the *quantity* and *quality* of data being recovered from pottery was rising. The number of individual features studied from each sample increased, and the range of values able to be assessed with measurable precision for each of these features also rose. The result has been the assembly of significantly larger sets of multivariate data. These require the use of computers both for storage and for the application of statistical procedures which allow us to begin to interpret the potentially numerous and complex relationships amongst such large numbers of variables.

Methods such as principal components analysis or cluster analysis are widely used by archaeologists for dealing with the numeric data generated by geochemical studies of ceramic fabrics. However, there still exist problems of a practical nature. One of these is, of course, simple access to statistically viable sets of samples, a persistent problem which is most familiar to those scholars choosing to work in foreign countries (countries whose exotic appeal is, sadly, directly proportional to their xenophobia and bureaucratic undergrowth). Another problem is the continuing need for awareness by archaeologists of the importance of using reliable physical data sets which have been created with a high degree of objectivity using reproducible standards. Only such carefully assembled data can help to minimise the degree of distortion inherent in reducing the features of a continuous process (*i.e.* handmade pottery) to discrete values. Both of these problems are permanent aspects of our work; but with a seasoned sense of humour in the first, and with a commitment to reducing the subjective element in the second, progress can be sustained.

There is, however, a third problem, less frequently acknowledged perhaps, but with comparable risks of distortion to research; it consists of the pragmatic realities, not to say xenophobia and bureaucratic undergrowth, in university Computer Centres: shortages of staff, software, and time for consultation and project design, and jargon-ridden language. To any outsider approaching with even a slight degree of trepidation these conditions can prove a fatal academic blow. To the more determined researcher they seem endless frustrations in what had appeared to be a clear path. What will be discussed here is the experience of just such a determined archaeologist, who had the audacity, or naïvety, to assume that statistical assessments of carefully collected data were no more trouble than the artistic arrangement of a similarity matrix or two, and who by serendipity found a new way around old obstacles.

7.2 The archaeological context

The archaeological work concerned doctoral research on Base Ring Ware, a distinctive Late Bronze Age pottery made in Cyprus and exported to large areas of the Near East and Egypt. The aim of the project was a comprehensive regional study of the ware's fabrics as they were found in Cyprus, for the purposes of material characterisation and provenance studies. A preliminary fabric typology was constructed from a large sherd sample of the ware, and consisted of six categories each of a TYPE and PASTE. Observations of macroscopic features related to pre-firing manufacturing and surface finish defined a sample's TYPE, while features of material composition and firing technology defined a sample's PASTE. Eighteen variables considered most consistent within a single vessel, and felt to be most easily reduced to numerical values with minimum loss of precision were used in a detailed study of a representative sub-sample of 230 Base Ring sherds. The range of possible values for each variable was defined and each value given a number code for purposes of creating a data matrix. This resulted in a data matrix of 4,140 cells overall.

Every effort was made to reduce the degree of subjectivity in the role of the observer by employing widely-accepted standards of measure. For example, the Munsell Soil Colour Chart was used to describe colours; fabric hardness was studied by means of a modified Moh's Test; and Visual Percentage Estimation Chart and Wentworth Grain-Size Scales, used in the study of sediments, were applied to inclusions. Technical terms used in description were defined in a glossary. Features which were found to vary greatly in degree over whole vessels (such as lime-spalling) were considered unreliable as criteria.

The resulting set of data included variables with three distinct metrics. **type** and **paste** assignments, **chronological label**, **degree of compaction**, **hardness** and **surface lustre** were ordinal. But the majority of variables were categorical; those of the presence-absence or yes-no variety: **subsurface horizon**, **polish or burnish**, **scraping**, **pre-fired paint application** and **evidence of double-firing**; and the multivalued categorical variables, **site name**, **firing horizon variety**, **vessel part** and **original vessel form**. Only **wall width** and **inclusion percentage** had the well-defined ratio metric.

These data were first subjected to two-way frequency tests, using BMDP1F to determine useful associations *between* specific variables (including significant negative associations). Such tests were able to show for example that fabric hardness, potentially a function of any of several factors in ceramics such as grain size, paste composition or compaction, rate, atmosphere or time of firing, was in the case of Base Ring Ware a function of paste composition and firing atmosphere rather than secondary manufacturing procedures like paste compaction. In addition several PASTE categories were shown to have important negative associations with particular TYPE categories. These tests were also able to identify individual technological features of the ware which varied in useful patterns over time and space. Such results proved extremely helpful in the overall characterisation of the ware and in discussions of distinctive groups of fabrics which were manufactured in different areas of the island over time.

The focus of this part of the research also included testing the quality of the hypothetical fabric groups set out in the preliminary typology. Therefore it was proposed to run clustering procedures on the data from the sherd subsample. Since each sample had been studied only as a numbered item with SITE NAME and CHRONOLOGICAL LABEL added later, the degree of objectivity in the study was felt sufficient to treat the archaeological samples as an unfamiliar ware. It was thereby hoped to arrive at meaningful groups of material unaffected by any pre-existing professional prejudices. Cluster analysis was felt to be the best procedure for assessing such a large and complex set of theoretically unfamiliar data in an objective and exploratory way.

It was hoped to be able to run the cluster analysis program both on the cases and then on the variables since the function of certain variables as *groups* of related values was important in defining the criteria for the fabric categories. It was also hoped that the clusters of cases would reflect the preliminary results from the frequency tests, with groups of samples showing regional and chronological patterns in technology. Further complexity was added to these proposals however by the requirement that the clustering procedure be able to accept the three *varieties* of variable (that is, categorical, ordinal and ratio) without significant statistical distortions. To this end it was decided to use Gower's coefficient of similarity to create a matrix for input to the clustering procedure. However, since Gower's is *not* one of the similarity measurements already in the BMDP program available at University College, the matrix first had to be constructed independently with the assistance and cooperation of a staff member in that College's Computer Centre. And it was at this point that certain obstacles of a pragmatic nature began to appear.

Developing a program to generate a Gower similarity matrix is, from a statistical point of view, quite simple. It is merely the repeated evaluation of the Gower similarity coefficient which is itself a simple weighted combination of conventional coefficients for variables with differing metrics. However, a 230 × 230 matrix is not a small beast. The time and space requirements for its manipulation succeeded in increasing the complexity of the program by several orders of magnitude. Nor was the system of linked GEC 4190 minicomputers at University College the best environment for this type of work, but of course they *were* accessible and familiar.

Eventually the similarity matrix was generated. Now, however, it turned out that the case and variable clustering programs (BMDP1M and BMDP2M) could not accept the matrix. This was because one of them will not accept an independently generated similarity matrix and the other could not cope with the size of matrix. Of course, given more time and knowledge these problems could have been anticipated or circumvented (there are other clustering programs around, for example), but, in our circumstances, the only viable option was to try the remaining, block clustering program (BMDP3M).

7.3 The 'block clustering' technique

This of course was the moment of serendipity (or as others might put it, the moment of desperate *post hoc* rationalisation!). What seemed at first sight to be a poor but acceptable alternative seemed to grow on reflection to become a more desirable option.

The peculiar merit of block clustering, as opposed to case or variable clustering or their combination, seems to be its ability to exclude what we here term 'relative noise'. Consider a data matrix. It may be homogeneous over the cases and the variables, that is, there is no grouping in either of these dimensions. In this case, clustering of any sort will (hopefully!)

fail to identify any significant structure within the matrix. Perhaps, however, the matrix is not homogeneous. It may be heterogeneous over the cases or the variables or both. Let us take the first case, that there is grouping in the case dimension but no grouping in the variable dimension. Now these case groups, or clusters, will be resolved, if at all, within a unified variable space, that is, we have values of each variable for each case. The particular variables have, presumably, been chosen precisely because it is expected that it will be possible to discover case clusters in terms of them; which is to say, the variables are supposedly *relevant*. For some sets of data this seems to be a reasonable assumption. For example, consider a set of body measurements (such as height and weight) taken from a sample of Europeans. There are no obvious *prima facie* grounds for suspecting that height, for example, *could* be a relevant discriminator for, say, the English but *could not* for, say, Spaniards. So, in this example, values of height can reasonably be assumed to be *equally* relevant in eliciting case structure.

But, for the ceramic data which was the subject of this research, we felt less confident that each variable would be equally relevant. It may be that the case clusters that we want to discriminate are distinguished by different ceramic technologies, for example, and precisely because these technologies *are* different, some variables may be highly relevant to characterising some of the technologies but irrelevant to characterising others. But, and this is the important point, in the data matrix we will still have values for these partially irrelevant variables. In a conventional *case* cluster analysis these subsets of irrelevant values would then constitute 'noise' relative to the process of distinguishing some clusters. Depending on the scale of the problem, the noise could either reduce the resolving power of the clustering algorithm or could lead to the identification of spurious clusters. It is in these circumstances that *block* clustering seems to offer advantages. On theoretical grounds it seems to offer a more discriminating approach, an approach better suited to excluding 'relative noise'. And the appropriateness of the technique seems to have been born out in practice.

The block clustering program was first run using the values for all 18 variables, and then rerun after eliminating the data from the variables SUBSURFACE, VESSEL PART and SITE NAME. These data were excluded for two reasons: the clusters from the first run appeared unaffected by the values for the first two variables, and SITE NAME was felt to wield an artificially important influence in defining regional patterns when the technological features were considered the primary criteria. The confidence levels for the second run clusters were significantly improved and the groups of sherds proved more meaningful in archaeological terms. Thus the importance of careful selection of discriminators was amply demonstrated.

In addition to the normal dendrogram produced for *case* associations the program also provided a dendrogram of variable clusters. Variables left unclustered or linked to only one other feature proved to be the least useful discriminators while *sets of values for variable pairs* which acted in concert were the most effective factors in creating meaningful groups of ceramic samples. This last point is very important for those archaeologists studying ceramic data by means of statistical procedures. Since ceramic features are most often the result of the interaction of *combinations* of particular manufacturing or compositional variables, it is vital that these relationships be well-established before their authority as discriminators of fabric groups can be accepted. The blanket application of clustering techniques to undigested sets of data will always result in clusters. Therefore the responsibility lies with the archaeologist to understand the nature and limitations of the data, the statistical nature of the variables and the potential for distortion in the chosen method of analysis.

In the case of the present example, several large block clusters emerged from the procedure which both reaffirmed the distinctions proposed for the Base Ring fabrics in the preliminary typology and which provided solid support for the regional and chronological patterns of development for the ware indicated in the tests of association. In subsequent laboratory examinations of these fabrics, geochemical, mineralogical and petrographic evidence provided further corroboration for the original macroscopic assessments and valuable data for material characterisation.

7.4 Conclusions

In the absence of computers and their software, by which archaeologists can both record and analyse large bodies of data, our increasingly sophisticated skills of excavation, retrieval and material analysis would be meaningless. Our ability to test hypotheses with numerous and complex sets of objective variables would also be seriously compromised, as would our understanding of the subtle interrelationships of these features of ancient technologies. Archaeological data is, inevitably, of a statistically mixed nature and the difficulties involved in resolving both the procedural as well as practical problems relating to the use of statistical techniques in archaeology are frequently exasperating and, without sympathetic guidance, often insurmountable. The only alternative to resourceful perseverance however is a relapse into the rampant nineteenth century subjectivity which characterised the discipline for too long. The museums are already filled with the spoils of eccentric characters in pith helmets—it is now more important to fill the computers with the less flamboyant spoils of research derived from these trophies.